

Het verborgen web ontsloten

Diepwebduiken

Gewone zoekmachines halen slechts een fractie tevoorschijn van de beschikbare informatie op het web, tenzij je ze extra slim gebruikt. Bovendien zijn er uitstekende indexen en speciale zoekmachines, waarmee we schijnbaar verborgen informatie alsnog aan de oppervlakte krijgen. DIRK SCHOOFS



WAT DOEN WE?
- INFORMATIE ZOEKEN DIE NIET TE VINDEN IS MET DE KLASSIEKE ZOEKMACHINES

WAARMEE?
- EEN INTERNETVERBINDING

HOELANG?
- 5 MINUTEN

MOEILIJKHEID?



Wie zich ernstig in een bepaalde materie verdiept, ondervindt al gauw dat hij niet zomaar snel alle internetwijsheid tevoorschijn kan googelen. De wijdverbreide veronderstelling dat zoekmachines niet verder geraken dan 30 tot 40 procent van de bestaande websites is zelfs nog te optimistisch. Volgens BrightPlanet, een bedrijf gespecialiseerd in zoektechnologie, is het braakliggend terrein voor zoekmachines veel groter dan we ooit konden vermoeden. Zo'n 2 miljard pagina's zou je kunnen bereiken met de traditionele zoekmachines (the surface web), maar er zijn 500 keer meer pagina's die zich niet rechtstreeks laten benaderen door de zoekmachines. Dat wil echter niet zeggen dat we ze niet kunnen vinden...

Hoe diep is het web?

Zoekmachines zoals Google en Yahoo! kunnen hun ding doen omdat de auteurs zelf hun werk aanmelden. Bovendien draaien er op hun servers zogenaamde crawlers of spiders, die hyperlink na hyperlink volgen. Hierdoor zijn de zoekmachines in staat om ook zelf webpagina's op te slaan in hun indexen. Het duurt 2 tot 6 maanden voor de conventionele zoekmachines nieuwkomers in hun indexen opnemen, en hiermee zijn we bij een eerste oorzaak van het onzichtbare web.

CompletePlanet™ The Deep Web Directory — Powered by BrightPlanet®
HOME | HELP / FAQ | CONTACT US | ABOUT COMPLETEPLANET | TO BRIGHTPLANET
Discover over 70,000+ searchable databases and specialty search engines.
Find databases relevant to: orchid — All of these words — Go!
ADVANCED SEARCH
All Documents — All Documents Within Home & Garden Tree
A comprehensive listing of dynamic searchable databases. Find databases with highly relevant documents that cannot be crawled or indexed by surface web search engines.
Related Databases 1-8 of 8 — Visit and submit a query to any of the relevant databases below. (Our DQM searches multiple databases simultaneously, and more – Click here)
1-8
1. Stencils from Stencil Ease
Summary: Bargain Basement Children's Stencils Country Stencils Disney Fish Stencils Floral Stencils Foliage Stencils Fruit Stencils Garden ... Stencils Holiday Inspirational Stencils Ivy Lettering Stencils Mural Stencils Nautical Stencils Neo Classical Stencils Southwest Stencils ... matches Search by Key Word **Orchid** Stencil 7.99 Palm Tree Stencil 38.00 Bamboo Stencil ... Mural Stencil Kit 71.91 for 9 stencil designs All Things Bright and Beautiful Complete Mural ... Stencil Kit 130.00 for 8 stencil designs Peaceful Valley Complete Mural Stencil Kit 120.00 ...
http://stencilease.com/db/index.cfm?catagory=mural
Relevance: Size: 24k Harvested: Fri Mar 05 12:51:45 CST 2004
Top of Page
2. eat2eat - Bangkok ...
Summary: is free to the user! SEARCH RESERVE FREQUENT DINING Learn how quickly you can
Ads by Google — Genealogy Search Engine — Discover the meaning and origin of your family name. www.FamilyEducation.com — Cluster Your Results

1

Eerst openen we de rubriek Home & Garden en daarna gebruiken we de zoekterm 'Orchid'.

Onder het onzichtbare of diepe web verstaan we informatie die wel degelijk bereikbaar is via het internet, maar die niet ontsloten wordt door de grote zoekmachines. Hier zijn tal van redenen voor. De crawlers van zoekmachines herkennen uitsluitend statische webpagina's. Dynamische pagina's, die hun actuele informatie genereren op vraag van de bezoeker, blijven verborgen voor de klassieke zoekmachine. Het hoofdbestanddeel van het onzichtbare web bestaat echter uit databases waar je pas toegang tot krijgt nadat je een gebruikersnaam en een wachtwoord hebt ingegeven. Ook heel wat pagina's van kranten en vakbladen zijn in het begin wel publiek toegankelijk, maar wil je ze enkele maanden later opnieuw lezen, dan krijg je plots geen toegang meer, terwijl die pagina's toch nog op de servers staan.
Hoe kan je toch informatie vinden op het onzichtbare web? Er zijn drie methoden: je hebt de speciale directories, de gespecialiseerde zoekmachines, of je kan alsnog de klassieke zoekmachines gebruiken als een tussenstap, om zo de databases te vinden die toegang geven tot het diepe web.

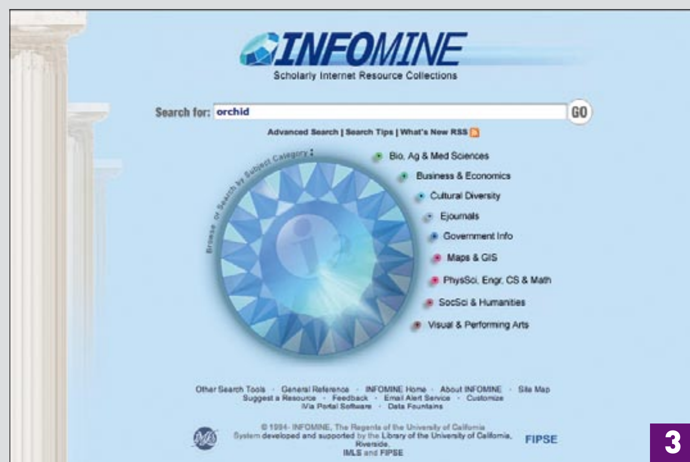
Speciale directories

CompletePlanet **www.completeplanet.com** is een site van BrightPlanet (zie afbeelding 1). De directory is in het Engels opgesteld. Hier spreek je 70.000 databanken en 11.000 gerelateerde zoekmachines aan. Ofwel zoek je via trefwoorden, ofwel gebruik je de overzichtelijke directorystructuur met 42 onderwerpen en honderden subrubrieken. Dit is een uitstekende bron voor onderzoekers en studenten, voor wie de tradi-

- Education (51)
- Emergency Services (527)
- Environmental Health (219)
- First Aid (24)
- Fitness (273) NEW!
- General Health (105)
- Health Administration (51)
- Health Care (291)
- Health Sciences (34)
- Hospitals and Medical Centers (44)
- Hygiene (17)
- Institutes (29)
- Job and Employment Resources@
- Pharmacy (1863) NEW!
- Procedures and Therapies (615) NEW!
- Professional Supplies and Services@
- Public Health and Safety (2464)
- Reference (81)
- Reproductive Health (771) NEW!
- Senior Health (97) NEW!
- Sexual Health (20)
- Teen Health (56)
- Traditional Medicine (152)
- Travel Health and Medicine (25)
- Web Directories (31)
- Weight Issues (277)

2

Alleen in de rubriek Health zitten er nog eens 31 web directories.

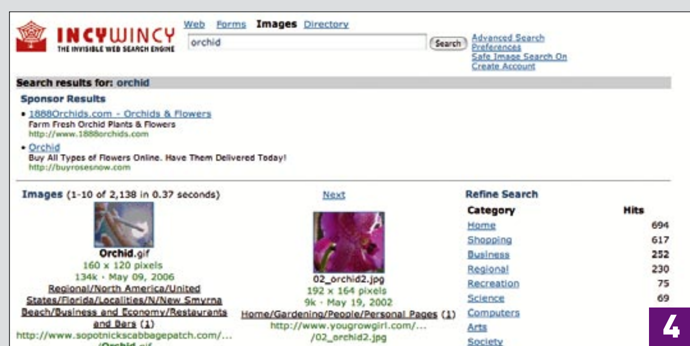


In Infomine staat de informatie gegroepeerd per vak.

tionele zoekmachines niet volstaan. Het nut van de ingebouwde zoekmachine op dit adres lijkt ons overroepen, tenzij je eerst naar de juiste subrubriek navigeert en pas dan het trefwoord ingeeft. Een andere site is Yahoo! Directory http://dir.yahoo.com. Surf er naar toe en klik door naar de juiste rubriek. Vervolgens krijg je per onderwerp een alfabetische categorielijst te zien. In iedere categorie vind je de rubriek **Web Directories** terug, die je leidt naar gegevensbanken die hun inhoud vaak verstoppen voor zoekmachines (zie afbeelding 2).
BUBL Information Service http://bubl.ac.uk is een uitstekende catalogus voor alle takken in het academisch onderwijs. Dit is een typische deep web-index van zorgvuldig geselecteerde bronnen. Ofwel gebruik je de geïntegreerde zoekmachine of de alfabetische index, ofwel raadpleeg je BUBL thematisch. De catalogus is gebaseerd op het Dewey Decimaal Classificatie-systeem dat in bibliotheken gebruikt wordt. Ook Infomine Scholarly Internet Resource Collections http://infomine.ucr.edu is een erg degelijke wetenschappelijke directory die toegang geeft tot meer dan 20.000 databanken, elektronische tijdschriften en websites van academisch niveau (zie afbeelding 3). In de zoekopties kan je de opdracht beperken volgens veldcategorie (auteur, titel, volledige tekst, ...) of volgens onderwerp. Ook hier zijn de zoekresultaten meestal van uitstekende kwaliteit. De organisatie gaat prat op haar "Expert-selected resources", ofwel het feit dat hun bronnen door experts werden geselecteerd. In het zoekresultaat lees je of de selectie automatisch tot stand kwam, of dat ze werd aangebracht door een bibliothecaris.

Gespecialiseerde zoekmachines

De speciale diepwebzoekmachines brengen in de meeste gevallen veel minder resultaten tevoorschijn dan de klassieke search engines. Ze moeten het meer hebben van de kwaliteit van hun resultaten dan van de kwantiteit.

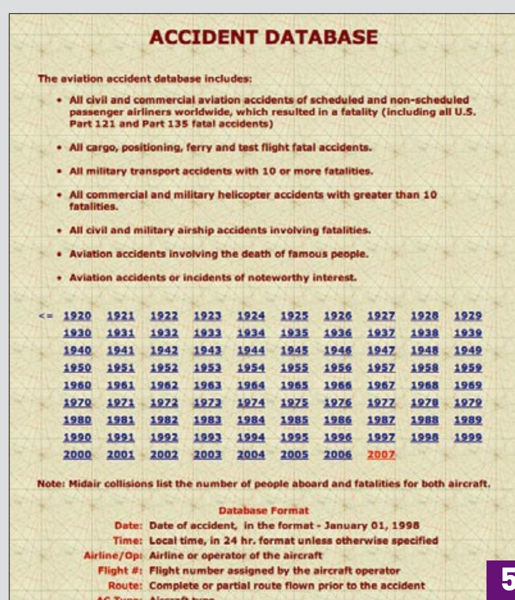


Je kan de zoekresultaten nog thematisch verfijnen.

IncyWincy www.incywincy.com is een veelzijdige zoekmachine, speciaal ontworpen voor het deep web (zie afbeelding 4). IncyWincy maakt gebruik van 150 miljoen geïndexeerde en doorzoekbare webpagina's en zoekmachines. IncyWincy heeft een eigen beeldenzoeker en ook een thematische directory.
GoshMe www.goshme.com zit nog steeds in een bètastadium. Gebruikers moeten zich registreren. Hier maak je simultaan gebruik van 2.520 deep web-zoekmachines en databanken. De resultaten worden opgedeeld in gespecialiseerde of algemene zoekmachines. Bovendien krijg je bij elke website enige toelichting over de betrokken organisatie of zoekmachine. Wie medische databanken wil aanspreken, moet Mamma health www.mammahealth.com onthouden. Dit is een typische 'deep web searcher' die tegelijk 8 medische gegevensbanken naslaat en als resultaat een lijst van artikels presenteert.

De klassieke zoekmachine als tussenstap

Zweer je toch bij de gewone zoekmachines, dan kan je ook die gebruiken om tot bij de poorten van het deep web te komen. Gebruik in dat geval enkele toevoegingen. Vat het onderwerp waarover je informatie zoekt in een paar termen samen en voeg daar woorden als 'database', 'data', 'dataset', 'archive', 'bibliography', 'index', 'register' of 'statistics' aan toe. Aanhalingstekens geven aan dat de woorden als een vaste combinatie worden gezocht. Wie dus met 'aviation accidents database' naar informatie over vliegtuigongevallen zoekt, komt inderdaad tot bij een database die alle fiches van vliegtuigongevallen bewaart (zie afbeelding 5).
Ook de toevoegingen 'intitle' en 'inanchor' brengen je naar de gegevensbanken. Wil je bijvoorbeeld alle titels zien van webpagina's die de naam Tertullianus bevatten, dan tik je 'intitle:tertullianus' (zonder de aanhalingstekens). Met 'inanchor' laat je Google alleen zoeken in de ankerverwijzingen van een html-pagina. Op die manier vind je koppelingen in webpagina's.



Verstandig zoeken leidt ons naar de gegevensbank van vliegtuigongevallen.

Wie eenmaal de binnenweggetjes naar het deep web kent, verhoogt zijn zoekpotentieel enorm. Net zoals de weg naar uw binnenzak, richting portefeuille, het hart van deze gids alweer sneller doet slaan. ♦